Statistical Learning, Syllable Processing, and Speech Production in Healthy Hearing and Hearing-Impaired Preschool Children: A Mismatch Negativity Study

Esther Studer-Eichenberger,[1] Felix Studer-Eichenberger,[2] and Thomas Koenig[3,4]

Objectives: The objectives of the present study were to investigate temporal/spectral sound-feature processing in preschool children (4 to 7 years old) with peripheral hearing loss compared with age-matched controls. The results verified the presence of statistical learning, which was diminished in children with hearing impairments (HIs), and elucidated possible perceptual mediators of speech production.

Design: Perception and production of the syllables /ba/, /da/, /ta/, and /na/ were recorded in 13 children with normal hearing and 13 children with HI. Perception was assessed physiologically through event-related potentials (ERPs) recorded by EEG in a multifeature mismatch negativity paradigm and behaviorally through a discrimination task. Temporal and spectral features of the ERPs during speech perception were analyzed, and speech production was quantitatively evaluated using speech motor maximum performance tasks.

Results: Proximal to stimulus onset, children with HI displayed a difference in map topography, indicating diminished statistical learning. In later ERP components, children with HI exhibited reduced amplitudes in the N2 and early parts of the late disciminative negativity components specifically, which are associated with temporal and spectral control mechanisms. Abnormalities of speech perception were only subtly reflected in speech production, as the lone difference found in speech production studies was a mild delay in regulating speech intensity.

Conclusions: In addition to previously reported deficits of sound-feature discriminations, the present study results reflect diminished statistical learning in children with HI, which plays an early and important, but so far neglected, role in phonological processing. Furthermore, the lack of corresponding behavioral abnormalities in speech production implies that impaired perceptual capacities do not necessarily translate into productive deficits.

Key words: Children, Cortical auditory–evoked potentials, Mismatch negativity, Peripheral hearing loss, Statistical learning, Temporal/spectral sound-feature processing

(Ear & Hearing 2015;XX;00–00)

INTRODUCTION

Humans distinguish speech sounds based on spectral and temporal properties, which are represented by segmental units such as laryngeal, place, and manner features (Hall 2007). Importantly, the maturation of this capacity during infancy crucially depends on exposure to speech sounds, suggestive of dedicated learning mechanisms (Kuhl et al. 2008). Among the major theoretical views of speech sound perception and categorization processes in children, the two most prominent are statistical learning (Aslin et al. 1999; Saffran 2003; Werker & Tees 2005) and the "Perceptual Assimilation Model" (Best 1995).

According to Thiessen et al. (2013), the mechanisms of statistical learning are rooted in conditional, distributional, and cue-based sensitivities to input information. The conditional aspect of statistical learning refers to transitional probabilities that indicate the predictive relationship between two events and is used to segment words, particularly in identifying prosodic information (Bertoncini 1993) and phonotactic structures (Saffran & Thiessen 2003), even in an unfamiliar language (Peluchi et al. 2009). The distributional aspect of statistical learning detects the tendencies within a categorical structure or prototypical characteristics of a given input. Through these distributional mechanisms, learners are sensitized to the frequency and variability of sets of presented elements (Thiessen et al. 2013), which plays a key role in categorizing speech sound features (Maye et al. 2002; Werker et al. 2007; Maye et al. 2008). Cue-based statistical learning is a mechanism for perceptual cue detection and weighting that recognizes the relationship between perceptual characteristics of the input and the units that organize the input (Thiessen et al. 2013). This is important for identifying acoustic cues to word boundaries (Thiessen & Saffran 2003) and speech sound perception (Nittrouer 1996). Among these facets of statistical learning, the distributional aspect is particularly pertinent to speech sound perception and categorization processes.

Another theoretical basis for the development of speech sound processing is the Perceptual Assimilation Model (Best 1995), which stresses the close connection between speech perception and production processes. This model emphasizes the importance of an integrated perceiving-acting system that is able to detect a broad range of possible articulatory gestures in speech processing. Articulatory gestures are composed of information extracted from all sensory modalities. Through an experience-based attunement to the constellations of gestures they are exposed to, infants discover which articulatory gestures are used in their native language and then categorize them into language-specific patterns. According to this theoretical framework, speech sound processing is based on perceptual abilities that are linked to speech production represented by articulatory gestures and involves increasing attunement (Best 1994, 1995) to speech sound categories over the course of language acquisition (Tyler et al. 2014).

Both distributional statistical learning and the Perceptual Assimilation Model implicitly represent different learning processes for perceiving speech sounds and creating speech sound categories. As a result, distinct training schemes are necessary to treat children with impaired speech sound perception, as is the case in children with peripheral hearing loss (HL) (Bamford

[1]Institute of Linguistics, University of Bern, Bern, Switzerland; [2]Institute and Department of Special Education, University of Fribourg, Fribourg, Switzerland; [3]Translational Research Center, University Hospital of Psychiatry, University of Bern, Bern, Switzerland; and [4]Center for Cognition, Learning, and Memory, University of Bern, Bern, Switzerland.

<zdoi; 10.1097/AUD.0000000000000197>

& Saunders 1985; Jerger 2007). To conduct ideally targeted auditory training for hearing-impaired (HI) children, empirical information about the relative contribution of the specific model mechanisms for speech sound perception and categorization learning is necessary. Based on the assumption that different learning mechanisms show distinct neuronal representations, collecting neurophysiological data of syllable processing will shed light on the mechanisms that underlie speech sound perception and categorization learning. This information can in turn be leveraged to facilitate well-directed auditory training.

Recording auditory event-related potentials (aERPs) is an EEG method that has been used to characterize patterns of brain activity associated with syllable processing over the course of development. A component of aERPs termed Mismatch negativity (MMN) is a useful electrophysiological tool for assessing the perception of speech sounds in children (Näätänen 2001). Specifically, MMN responses are an effective index of auditory change detection processing in which the incoming stimulus is differentiated from the sensory memory trace of the preceding auditory inputs, which encodes regularities of successive stimulus events. Analysis of MMN thereby makes it possible to determine auditory discrimination accuracy and to assess abnormal auditory perception processing (Näätänen et al. 2012).

Previous studies investigating event-related potentials (ERPs) during syllable processing in typically developing children between 4 and 10 years of age parsed the syllable-evoked brain responses into several components. For example, the P1 component of the EEG waveform response to an external speech stimulus (Maurer et al. 2003; Ceponiene et al. 2005; Ceponiene et al. 2008; Sussman et al. 2008) has been divided into an early bilateral frontal (80 to 120 msec) subcomponent followed by a more central (150 to 180 msec) subcomponent (Ceponiene et al. 2005). Several studies (Gilley et al. 2005; Wunderlich et al. 2006) provided evidence of distinguishable P1 and P2 components with peaks at central electrode locations, but unsystematic alterations in peak latency with increasing age (Wunderlich et al. 2006); however, others could not find a P2 component (Ceponiene et al. 2005; Ceponiene et al. 2008; Sussman et al. 2008). Another EEG component, the so-called N1, was observed in the Cz and T3 electrodes in a time window around 160 msec (Wunderlich et al. 2006) depending on stimulation rate and age; detectability is increased in children older than 10 years (Gilley et al. 2005). Further components have also been identified, such as the N2 with a maximal frontocentrally peak around 300 msec (Maurer et al. 2003; Ceponiene et al. 2005). In previous studies of adult subjects, the N2 has sometimes been subdivided into three components: N2a, N2b, and N2c. N2a is a relatively attention-independent component, which corresponds to the MMN and is maximal at frontocentral electrodes (a time window from 150 to 250 msec) (Kujala et al. 2001). N2b is a more attention-dependent and anticipatory component with a frontocentral maximum (a time frame from 250 to 300 msec). N2c is the categorizing component, which shows a central negativity that is maximal at more posterior electrodes (slightly later than 300 msec) (Näätänen 1987). However, little is known about these components in children aged 4 to 10. In contrast, the N4 component has been detected in children with a maximum at frontocentral electrodes in a slightly later time frame (370 to 450 msec) (Ceponiene et al. 2005).

The functional implications of the P1/P2 complex have been suggested to represent preperceptual processing of acoustic sound features (Ceponiene et al. 2005), such as feature detection and identification of spectral (Ceponiene et al. 2008) and temporal differences (King et al. 2008), whereas distinct P2-related processes have been proposed to play an additional role in automatic attention mechanisms (Ceponiene et al. 2005). Later components, such as N2 and N4, are thought to be important for integrative encoding of sound content features (Ceponiene et al. 2005; Ceponiene et al. 2008), and it has been noted that N2 latency and amplitude changes were influenced by temporal differences in voice onset time (VOT) (King et al. 2008). In terms of N2 subcomponents, the function of N2a has been ascribed to passive discrimination processes, whereas the N2b is represented in more active discrimination processes or stimulus-specific anticipatory processes. The N2c component is associated with categorization processes to various auditory stimuli (Näätänen 1987).

Investigations in speech sound acquisition in children with peripheral HL revealed disabilities in spectral and temporal information processing (Bamford & Saunders 1985; Jerger 2007). Depending on the degree of hearing dysfunction, abnormal processing of VOT differences have been observed, indicating impaired perception of the temporal aspects of speech sounds (Jerger 2007). Abnormal perception of spectral information has been described for the integration of mid-frequency and widely separated frequency bands; these findings are particularly salient for place of articulation differences observed in subjects with HL (Grant et al. 2007). In addition, dysfunctions in auditory temporal ordering of language-specific stimuli have been observed to depend on the degree of HL (Koravand & Jutras 2013). Furthermore, children with impaired hearing have impairments in motor and visual sequential learning, suggesting that reduced auditory perception has an impact on cross-modal encoding and learning, and the manipulation of sequences (Conway et al. 2009).

Furthermore, children with HI were shown to process speech sounds in a more controlled rather than automatic manner (Jerger 2007), and they presented abnormal temporal speech behaviors particularly related to vocal fold articulation (Paterson 1994; Higgins et al. 2005). Findings from Higgins et al. (2005) indicated that most of the produced speech was close to normal apart from aberrant VOT values, which were more severe in female subjects with unbalanced hearing abilities in both ears. Regarding perseveratory coarticulation, a similar precision between children with normal hearing (NH) and HI was found for consonantal duration (Baum & Waldstein 1991). Stiles et al. (2012) reported that compared with NH control subjects, children who use hearing aids demonstrated slower articulation rates in sentence repetition, which is related to verbal short-term memory processes.

Neurophysiological studies in children with cochlear implants confirmed the appearance of the P1 and N2 components in syllable processing (Singh et al. 2004). Recently, children with peripheral HL, who were fitted with hearing aids, were evaluated in a MMN study (Koravand et al. 2013). These authors recorded ERP responses with verbal and nonverbal acoustic transformation of /da/ and /ba/ stimuli, and pure tones (1 and 2 kHz), in 9- to 12-year-old French-speaking children. No significant differences in the P1 latency and amplitude or the latency of the N2 component between participants with healthy hearing and HI were observed. In contrast, a significant

group effect indicated reduced amplitude of the N2 component in children with HL. Hence, the authors concluded that the earlier sensory stages of syllable processing did not differ between children with hearing aids and children with NH.

In comparison with the extent of behavioral differences between HI and NH children, the neurophysiological group differences in the aforementioned studies were small. This may be because several important aspects of speech sound perception have not yet been accounted for. First, there are no available neurophysiological data describing syllable processing in children with hearing aids younger than 9 years old, even though most hearing-specific interventions begin during preschool age. Therefore, characterizing ERP responses of preschool children is necessary to decipher the sequences of processing steps, which would in turn yield information about what aspects of auditory stimulation are adequate for well-directed auditory training. Second, prior ERP studies were based on single channel analysis. We presume that multichannel analysis would significantly increase the sensitivity of ERP group analysis. Third, most of the previous studies of speech sound perception were focused on spectral features, such as the /ba/ versus /da/ distinction, while the contribution of temporal features has received little attention. However, behavioral results in children with HL have demonstrated a lack of temporal perception (Jerger 2007), indicating that corresponding neurophysiological data would enrich our understanding of speech sound perception mechanisms. Fourth, according to speech sound acquisition theories (Aslin et al. 1999; Werker & Tees 2005), an important underlying implicit learning mechanism is the analysis of distributional information, which plays a key function in speech sound perception and categorization processes. Thus, our assumption is that the effects of distributional statistical learning should also be reflected in neuronal responses that arise during syllable processing. To our knowledge, there has yet to be an ERP study elucidating distributional statistical learning mechanisms in preschool children with and without HL. Finally, although the relationships between syllable-evoked ERP responses, speech perception, and reaction time (RT) were measured in adults with HL (Oates et al. 2002; Korczak et al. 2005), little is known about these relationships in syllable production in children with HL. Such analyses are important to test if perception and production mechanisms of articulatory gestures are connected, as predicted by the Perceptual Assimilation Model (Best 1995).

The aims of the present study were to use integrated multichannel ERP analysis and receptive and expressive behavioral tasks in an investigation of how temporal and spectral speech sound features are processed in normally developing 4- to 7-year-old children compared with age-matched children with peripheral HL. Tasks were designed to test for indications of distributional statistical learning in behavioral tasks or ERP responses, which allowed us to clarify the nature of statistical learning mechanisms. Finally, to assay for evidence of an integrated perception–action system as in the Perceptual Assimilation Model of Best (1995), we explored whether there is a connection between neurophysiological processes of speech sound perception and the production of speech sounds.

METHODS

Subjects

The recruitment of the HI participants has taken place in a center for hearing and language services, whereas the NH subjects were recruited via personal contacts with their parents. Seventeen healthy hearing, typically developing children (nine girls and eight boys; age, 4 years 2 months to 7 years 5 months; mean age = 6.07 years; SD = 0.90 years) and 13 children with HI (eight girls and five boys; age, 4 years 0 months to 7 years 10 months; mean age = 6.55 years; SD = 1.09 years) participated in this study. Two children with HI dropped out due to a neurological impairment and a complete refusal to undergo procedures necessary for neurophysiological measurements. The remaining 30 subjects were Caucasian, monolingual native German or Swiss German speakers from working- and middle-class families in the region of Bern. None of them was a native sign language speaker, but two of 30 participants were exposed to sign language 2 hr a week. All children had normal speech motor skills as assessed by maximum performance tasks (MPTs) (Thoonen et al. 1996; Thoonen et al. 1999) and normal vision. All subjects were educated in the regular school system and none had a history of neurological development or behavioral problems. Children with NH had a sensitivity loss < 20 dB HL in both ears. Children with a peripheral HL had > 30 dB HL in at least one ear, as determined by pure-tone audiograms (125 to 8000 Hz). Furthermore, children with HI associated with meningitis or cytomegalovirus infection were excluded from the study. These diseases can cause damage to specific neural structures that are important for speech sound processing. Detailed subject and demographic information is provided in Tables 1 and 2.

Of the 30 subjects, 13 children each participated in the HI group and NH group with no differences in age ($t = 1.21$, degrees of freedom [df] = 24, $p = 0.2380$) or sex (Chi-square = 0.1548, df = 1, $p = 0.6940$). After explaining the study to each participant and his parents, written informed consent was obtained. As compensation, each child received a small present after the session. This study was approved by the local Ethics Committee of Canton Bern, Switzerland.

Stimuli

To meet the particular needs of this study, the selected stimuli should possess the following properties: there should be systematic differences in (1) spectral, (2) temporal, (3) spectrotemporal features of the speech sounds; (4) the range of frequency of occurrence of syllable onset should be as similar as possible; and (5) the distribution of speech sounds should

TABLE 1. Detailed information of subjects with normal hearing

Code	Age	Sex
K10	4 yr 2 mo	Female
K12	4 yr 9 mo	Male
K15	5 yr 1 mo	Female
K5	6 yr 1 mo	Male
K17	6 yr 3 mo	Female
K14	6 yr 3 mo	Male
K6	6 yr 4 mo	Female
K8	6 yr 6 mo	Female
K16	6 yr 6 mo	Female
K4	6 yr 8 mo	Male
K9	6 yr 9 mo	Male
K7	6 yr 9 mo	Male
K1	7 yr 5 mo	Male

TABLE 2. Detailed information of subjects with peripheral hearing loss

Code	Age	Sex	HI	AA	Etiology	Ear	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
U5	4 yr 0 mo	F	B	2 yr 0 mo	Unknown	L	60 dB	60 dB	60 dB	60 dB	60 dB	50 dB	50 dB
						R	60 dB	60 dB	60 dB	60 dB	70 dB	80 dB	80 dB
U13	4 yr 9 mo	F	B	4 yr 5 mo	Chemo induced	L	NA	20 dB	25 dB	15 dB	55 dB	95 dB	75 dB
						R	NA	35 dB	30 dB	20 dB	65 dB	90 dB	80 dB
U12	5 yr 11 mo	M	B	1 yr 2 mo	Unknown	L	NA	90 dB	105 dB	90 dB	85 dB	80 dB	65 dB
						R	NA	85 dB	100 dB	75 dB	70 dB	85 dB	70 dB
U15	6 yr 5 mo	F	B	UA	Unknown	L	20 dB	20 dB	20 dB	40 dB	30 dB	30 dB	30 dB
						R	20 dB	20 dB	20 dB	40 dB	30 dB	30 dB	30 dB
U1	6 yr 9 mo	F	MO	5 yr 8 mo	Unknown	L	NA	NA	30 dB	50 dB	65 dB	70 dB	NA
U2	6 yr 10 mo	M	B	5 yr 8 mo	Unknown	L	NA	100 dB	110 dB	100 dB	90 dB	75 dB	70 dB
						R	50 dB	50 dB	45 dB	25 dB	15 dB	25 dB	50 dB
U11	7 yr 0 mo	M	B	6 yr 5 mo	Unknown	L	30 dB	30 dB	25 dB	25 dB	15 dB	5 dB	NA
						R	30 dB	30 dB	25 dB	25 dB	20 dB	5 dB	NA
U10	7 yr 0 mo	F	B	4 yr 4 mo	Hereditary	L	NA	40 dB	55 dB	60 dB	60 dB	60 dB	55 dB
						R	NA	35 dB	45 dB	50 dB	60 dB	55 dB	50 dB
U9	7 yr 3 mo	F	MO	6 yr 8 mo	Unknown	R	80 dB	80 dB	80 dB	80 dB	70 dB	65 dB	25 dB
U14	7 yr 3 mo	F	B	6 yr 10 mo	Hereditary	L	NA	10 dB	30 dB	25 dB	45 dB	60 dB	60 dB
						R	NA	20 dB	20 dB	20 dB	20 dB	40 dB	50 dB
U8	7 yr 5 mo	F	B	4 yr 10 mo	Unknown	L	NA	20 dB	20 dB	45 dB	65 dB	65 dB	65 dB
						R	NA	15 dB	15 dB	20 dB	65 dB	55 dB	50 dB
U7	7 yr 5 mo	M	B	4 yr 10 mo	Unknown	L	NA	45 dB	50 dB	50 dB	50 dB	65 dB	70 dB
						R	NA	70 dB	70 dB	70 dB	70 dB	70 dB	70 dB
U3	7 yr 10 mo	M	MO	UA	Microty	R	NA	70 dB	70 dB	70 dB	65 dB	65 dB	55 dB

AA, age of amplification; B, binaural; F, female; HI, hearing impairment; L, left ear; M, male; MO, monaural; NA, not available; R, right ear; UA, unamplified.

cover low-, mid-, and high-frequency bands. To assess a spectral difference, our stimulus required a difference of place of articulation representing a spatial distinction within the same phoneme category. For assess a temporal difference, the VOT is a particularly important feature because the distinction between two phonemes exists only in the different onsets of the vocal cords vibrations. Features of the spectrotemporal aspect can be selected by a phoneme distinction through a different manner of articulation. In other words, the difference between two phonemes is revealed in dissimilar phoneme categories and phoneme durations (Stevens 1998; Pompino-Marschall 2009; Reetz & Jongman 2009; Johnson 2012). Based on various German studies (Best 2004/2005; Aichert et al. 2005) and a Swiss German study (Siebenhaar, Reference Note 1), it is known that the frequency of syllable onset of several plosives and the alveolar nasal are within the first nine ranges. Voiceless stops are within high-frequency ranges, while voiced plosives are related to mid-frequency band and nasals to low-frequency bands (Stevens 1998; Reetz & Jongman 2009). Therefore, we chose the syllables /ba/, /da/, /na/, and /ta/. As a result, the /da/ stimulus had a VOT distinction to /ta/ covering the high-frequency band, while /da/ had a spectral difference to /ba/ represented by a transition between the second and the third formant (mid-frequency band). The spectrotemporal distinction was manifest between /da/ and /na/ due to a variant phoneme classes (plosive versus nasal), as there is considerably longer phoneme duration for /na/ than /da/ and the nasal was assigned to the low-frequency band.

Stimuli were recorded by a microphone (Sennheiser Back-Electret MPR-65) with a sampling rate of 8000 Hz and arranged using PRAAT 5.1 speech analysis software (Boersma & Weenik 2009). The glottal waveform was extracted from the natural utterance of a female Swiss German speaker pronouncing the syllables with the same fundamental frequency (F0 180). Plosive speech sounds (/b/, /d/, and /t/) were cut to a duration of 35 msec (Pompino-Marschall 2009), whereas the nasal speech sound (/n/) had a length of 75 msec (Siebenhaar, Reference Note 1). Even though the speech sounds were generated from a natural voice, the syllables were semisynthetic products. The same vowel sound /a/ was added to each consonant, resulting in syllables /ba/, /da/, /ta/, and /na/. Detailed information about formant frequencies is provided in Table 3. All syllables were matched for fundamental frequency (F0 180 Hz), duration (300 msec), and intensity (70 dB). Stimuli were presented in a sound-attenuated chamber through a loudspeaker located 1.5 m in front of the child's head at an intensity of 75 dB SPL.

Tasks

To meet the aims of this study, we carried out the following three procedures: (1) a behavioral syllable discrimination task; (2) a phoneme and syllable production evaluation assessed by a modified version of the MPTs (Thoonen et al. 1996; Thoonen et al. 1999); and (3) neurophysiological measurements. Data regarding syllable discrimination were collected to clarify whether there were significant differences in

TABLE 3. Formant frequencies (F1, F2, F3, and F4) for the selected syllables (/ba/, /da/, /ta/, and /na/)

Formant Frequencies	Syllable /ba/	Syllable /da/	Syllable /ta/	Syllable /na/
F1	470 Hz	540 Hz	560 Hz	280 Hz
F2	955 Hz	634 Hz	1180 Hz	1190 Hz
F3	1600 Hz	1775 Hz	1700 Hz	2100 Hz
F4	2880 Hz	2935 Hz	2870 Hz	2940 Hz

discrimination accuracy between the NH and HI groups and between age groups. RTs were measured to rule out attentional confounders. Variations in the frequency of occurrence of syllable pairs were presented to assess the presence of a distributional statistical learning mechanisms. The modified MPTs were implemented to sort out participants with motor speech disorders and to investigate differences between the following variables: NH and HI groups, sex, hearing function, age, and duration and intensity of consonantal speech products, as well as the perseveratory precision of the consonants. The perseveratory precision of consonantal duration and intensity is of particularly importance for verifying whether there was a potential connection between speech sound perception and production as assumed in the Perceptual Assimilation Model (Best 1994, 1995). Electrophysiological measurements were performed to obtain a neurophysiological basis of speech sound processing of spectral, temporal, and spectrotemporal features, while also clarifying whether mechanisms of statistical learning were present. We also determined whether ERPs covariate with behavioral measurements and maturational effects.

The children were first asked to discriminate 40 syllable pairs, after successful completion of a short sample trial. They were informed which buttons they should press to indicate similar versus different syllable pairs and were instructed to press the corresponding button as quickly as possible after deciding on the type of stimulus. Syllable pairs comprised 20 identical and 20 dissimilar pairs with variant frequency of occurrence made up of five dissimilar pairs of items with difference in VOT /da:ta/, five with different place of articulation /da:ba/, five with unlike manner of articulation /da:na/, and five with infrequent syllable combinations, for example /na:ba/, all of which had to be differentiated by the subject.

After this, the MPTs (Thoonen et al. 1996) were performed. The children were instructed to generate speech sounds (/a/, /m/, /s/, and /f/), monosyllables (/pa/, /ta/, /ka/, /ba/, /da/, and /na/), and trisyllables (/pataka/, /katapa/, and /badana/) and were shown an example of what to do by the experimenter so that they could hear and see all speech sound movements. They were instructed to hold a given speech sound as long as they could or to repeat continuously a set of syllables as quickly as they could. Each task was measured from the beginning until the end of breath. All participants could repeat each trial three times, if the utterances were not correctly completed.

Third, neurophysiological measurements were implemented using a multifeature paradigm (Optimum 1) (Näätänen et al. 2004). The standard stimulus /da/ alternated with three deviant stimuli /ba/, /ta/, and /na/. Each sequence started with 20 standard stimuli. Altogether 1140 stimuli were presented to each subject, subdivided into 600 standard stimuli and 540 deviant stimuli (i.e., each deviant 180 times). The stimulus-onset asynchrony was 1000 msec with an interstimulus interval of 700 msec. The EEG recordings were repeated in three identical trials. Each sequence lasted 6 min and 20 sec, resulting in EEG registrations of approximately 20 min. During the EEG recording, the children watched a silent movie without human mouth movements. They were told to ignore the auditory input and to keep their attention on the movie. All tasks were completed in a set order for all children. The testing duration lasted 2.5 hr including behavioral tasks, EEG recordings, and breaks for rest.

Data Recordings and Data Analysis

Behavioral Tasks • The behavioral discrimination task was recorded and analyzed by the computer-based program E-prime (Schneider et al. 2001). Syllable discrimination accuracy was evaluated by descriptive statistics to quantify successful hits for all five distinct feature categories for the NH and the HI groups. To assess statistical significance, *t* tests were applied to evaluate age effects in both groups (younger, 4 years 0 months to 5 years 11 months; older: 6 years 0 months to 7 years 11 months). Discrimination accuracy was correlated to the degree of HL in the HI group according to the Council on Physical Therapy of the American Medical Association (CPT-AMA 1942) to evaluate whether the number of correct results depended on the severity of HI. Furthermore, repeated-measures analyses of variance (ANOVAs) for each syllable pair category were calculated between groups (HI and NH) and age groups. In addition, binomial tests of the NH and HI groups and both age groups were calculated to judge whether or not the amount of accurate hits went beyond chance level. To control for alternations in attentional processes, trends in RTs of each stimulus type (syllable pairs with spectral, temporal, or spectrotemporal distinctions, as well as infrequent and identical syllable pairs) and of all items for each subject over time were calculated using a general linear model. Some children needed to be repeatedly reminded to press the corresponding button and were therefore excluded from data analysis with respect to RT.

Speech production data were stored and processed using PRAAT 5.1 (Boersma & Weenik 2009). The first successful trial of each type of syllable was segmented into typical consonant waveforms. For voiced plosives, these segments were defined as a closure followed by a burst and an irregular perturbation before regular vocal waveforms appeared, while the voiceless stop showed an interval between burst and onset of vocal vibrations without irregular perturbations. For the nasals, the waveform was characterized through a smaller amplitude in the vocal fold pulses than in the subsequent regular vocal waveforms (Ladefoged 2003). To eliminate disruptive influences of higher speech motor variability at speech onset and to sustain short-term memory capacity, we evaluated the third through the eighth syllabic product. From these consonantal segments, we extracted the duration (msec) and the mean intensity (dB). A one-way repeated-measures ANOVA with consonantal duration and intensity as dependent variables and either group (NH and HI), sex, hearing function (monaural and binaural), age (younger, 4 years 0 months to 5 years 11 months; older, 6 years 0 months to 7 years 11 months), or consonants (/p/, /t/, /k/, /b/, /d/, and /n/) as independent variables was performed. A one-way repeated-measures ANOVA with standard deviation of consonantal duration and intensity was computed to evaluate the perseveratory precision between both groups. A *t* test was applied to examine whether there was a speech rate effect. The maximal amount of articulated monosyllables per second obtained from the MPTs (Thoonen et al. 1996) of each child represented the maximal speech rate. The values of maximal speech rate were correlated with the degree of HL measured by the CPT-AMA (1942) in the HI group to test the dependence of speech rate on the degree of HL. The level of significance for all tests of behavioral data was fixed at a p value of at least 0.05. All calculations were carried out using the statistical program R (R Development Core Team 2010).

Electrophysiological Measurements • The EEG data were collected with a Nihon Kohden Neurofax 1100 system at a sampling rate of 500 Hz (front-end band pass: 0.1 to 100 Hz) with 21 Ag/AgCl electrodes following the International 10 to 20 system (Jasper 1958) and two additional electrodes in the position of Fpz and Oz. The ground electrode was attached left paramedial to the electrodes Cz and Pz. C3 and C4 were used as reference electrodes. Eye movements were registered by means of two electrodes placed below the eyes. The impedance of each electrode was less than 10 kΩ. The EEG data were collected and stored on a personal computer for further processing and analysis utilizing Brain Vision Analyzer software 1.05 (Brain Products GmbH, Gilching, Germany). Off-line, biological, and electrical artifacts were removed using an independent component analysis to correct for eye movements (Jung et al. 2000), augmented by manual removal of visually detected residual artifacts. In addition, the initial 20 standard stimuli of each trial were removed. The data were segmented (100 msec before and 600 msec after the stimulus onset), band-pass filtered (0.1 to 30 Hz, and 48 dB/octave), recalculated against the average reference and within stimulus condition and each group. Group and condition-wise grand-mean ERPs were also computed.

The aim of the analysis was to discretely investigate the effects of topography, indicating differences in source localization, global amplitude representing the strength of activation, and latency specifying the processing speed. The analysis of topography was based on topographical analysis of variance (TANOVA) (Strik et al. 1998) with syllable and group conditions as repeated measures. Where the TANOVA indicated significant effects of map topographies, post hoc comparisons using *t* maps were carried out. To correlate individual traits with topographical changes, TANCOVAs (Koenig et al. 2008) were implemented. These individual traits included age and the maximal speech rate. Of the latter, the median was calculated to divide the expressive data into a group with higher and lower speech rates. The analysis of the global amplitude was carried out by calculating the global field power (GFP) (Lehmann & Skrandies 1980). Differences in timing of the microstates were statistically analyzed to determine the latency from stimulus presentation and to define their onset and offset and their duration. These are stable configurations of potential topographies for a certain period until they switch suddenly to another invariable configuration. Only a few are required to explain a great proportion of a data set (Michel et al. 2009).

Before the statistical tests were executed, the variances of the scalp fields across sensors were normalized for unit GFP. Significance level was always fixed at *p* value thresholds of 0.05. These calculations are based on randomization statistics and were carried out by the computer program RAGU (Randomization Graphical User interface) (Koenig et al. 2011).

RESULTS

Speech Perception Effects

Behavioral Results of Discrimination Accuracy of Syllable Pairs in NH Subjects • The NH group achieved a spectral discrimination accuracy of 27.69%, and the *t* test showed no significant age effect ($t = -0.0623$, df = 11, $p = 0.9514$). With respect to the temporal discrimination, they differentiated 20% of all VOT distinctions correctly. No age effect ($t = -0.6420$, df = 11, $p = 0.5341$) could be detected in the *t* test. Spectrotemporal feature discrimination was implemented with an accuracy of 66.15%, and no age effect ($t = -0.9959$, df = 11, $p = 0.3407$) appeared in the *t* test. Infrequent syllable pairs reached a discrimination accuracy of 41.53%, and a *t* test indicated there was no age effect ($t = -0.5243$, df = 11, $p = 0.6105$). Regarding syllable discrimination accuracy of identical syllable pairs, NH children realized 88.85% correct, and the *t* test did not reveal an age effect ($t = -1.7067$, df = 11, $p = 0.1159$). These results indicated that independent of age, there was an increasing degree of difficulty in discriminating features of same, spectrotemporal, infrequent, spectral, and temporal of syllable pair differences.

Behavioral Results of Discrimination Accuracy of Syllable Pairs in HI Subjects • The HI group correctly discriminated 23.07% of all spectral distinctions. No age effect could be observed in the *t* test ($t = 0.7389$, df = 11, $p = 0.4755$). Regarding temporal differences, 21.53% of the /da/ and /ta/ syllable pairs were discriminated accurately. The *t* test revealed no age effect ($t = 0.3694$, df = 11, $p = 0.7188$). The HI group perceived 64.61% of spectrotemporal differences correctly. The *t* test showed no age effect ($t = -0.5911$, df = 11, $p = 0.5664$). Of all infrequent syllable pairs, they differentiated 44.61% accurately. A *t* test demonstrated no age effect ($t = 0.1424$, df = 11, $p = 0.8894$). Regarding syllable discrimination accuracy of identical syllable pairs, the HI group attained 79.23%, and the *t* test revealed no age effect in the HI group ($t = -1.5017$, df = 11, $p = 0.1613$). These results led us to conclude that independent of age, children with HI show an increase of degree of difficulty with respect to same, spectrotemporal, infrequent, spectral, and temporal feature distinction in syllable pairs. The correlation between the amount of successful hits and the degree of HL was not significant ($r = -0.2604832$, $p = 0.3900$), indicating that the discrimination accuracy had not depended on the degree of HL.

Behavioral Results of Group Comparisons in Discrimination Accuracy of Syllable Pairs Between NH and HI Subjects • The ANOVA did not reveal a significant NH and HI group effect [$F(1, 22) = 0.15$, $p = 0.6993$] or group by age interaction [$F(1, 22) = 0.27$, $p = 0.6056$] concerning /da/ and /ba/ syllable discrimination and no NH and HI group effect [$F(1, 22) = 0.03$, $p = 0.8728$] with no group by age interaction [$F(1, 22) = 0.46$, $p = 0.5034$] between the syllable /da/ and /ta/. Further ANOVA results indicated that neither a significant NH and HI group effect [$F(1, 22) = 0.01$, $p = 0.9191$] nor a group by age interaction [$F(1, 22) = 0.09$, $p = 0.7669$] could be detected with respect to /da/ and /na/ syllable discrimination. The discrimination accuracy of infrequent syllable pairs revealed neither a group effect [$F(1, 22) = 0.07$, $p = 0.7942$] nor a group by age interaction [$F(1, 22) = 0.23$, $p = 0.6345$] in the ANOVA data. Neither a group effect [$F(1, 22) = 2.64$, $p = 0.1183$] nor a group by age interaction [$F(1, 22) = 0.25$, $p = 0.6231$] could be detected in identical syllable pairs. The binomial test revealed that children of NH and HI groups and in both age groups discriminated syllable pairs at chance level (NH younger $p = 0.5500$, NH older $p = 0.6650$, HI younger $p = 0.5250$, HI older $p = 0.6075$). In summary, children with HI displayed slightly diminished discrimination accuracy compared with the NH children with the exception of temporal and infrequent syllable pair distinction. In addition, there were no significant differences within hearing groups or age groups. It is possible that children with HI differentiate infrequent syllable pairs better than NH because their perceptive abilities were not sensitive enough to detect a difference in the frequency of occurrence. The behavioral results of discrimination accuracy are listed in the Table 4.

TABLE 4. Behavioral results of discrimination accuracy of syllable pairs in normal-hearing and hearing-impaired subjects and group comparisons

Speech Perception Effects	Normal-Hearing Subjects	Hearing-Impaired Subjects	Group Comparisons
Discrimination accuracy			
Spectral	27.69%	23.07%	—
Temporal	20%	21.53%	—
Spectrotemporal	66.15%	64.61%	—
Infrequent	41.53%	44.61%	—
Same	88.85%	79.23%	—
Age effects in syllable discrimination accuracy			
Spectral	$p = 0.9514$	$p = 0.4755$	$p = 0.6993$
Group by age interactions			$p = 0.6056$
Temporal	$p = 0.5341$	$p = 0.7188$	$p = 0.8728$
Group by age interactions			$p = 0.5034$
Spectrotemporal	$p = 0.3407$	$p = 0.5664$	$p = 0.9191$
Group by age interactions			$p = 0.7669$
Infrequent	$p = 0.6105$	$p = 0.8894$	$p = 0.7942$
Group by age interactions			$p = 0.6345$
Same	$p = 0.1159$	$p = 0.1613$	$p = 0.1183$
Group by age interactions			$p = 0.6231$

Behavioral Results of RT in NH and HI Subjects • None of the children showed an increase in RT over all 40 syllable pairs over time, indicating that there were no relevant attentional confounders in the behavioral data. However, 2 of 26 children showed a significant decrease of RT over all 40 syllable pairs over time.

Confirmatory Analysis in Physiological Correlates of Syllable Processing

The microstate analysis of the ERPs of the NH subjects revealed a series of components that were similar to descriptions in previous reports (Maurer et al. 2003; Ceponiene et al. 2005; Gilley et al. 2005; Wunderlich et al. 2006; Ceponiene et al. 2008; Sussman et al. 2008), while other components, such as the N1, were not detected. This might be explained by the fact that primary auditory-evoked potentials undergo substantial changes until the age of about 16 years (Mahajan & McArthur 2012), so that our data are not comparable to older subjects examined in other studies (Gilley et al. 2005; Wunderlich et al. 2006). The ERP microstate topographies and dynamics are shown in Figure 1A. Only significant time windows longer than 10 msec were interpreted for the analysis described in the following sections (Murray et al. 2008).

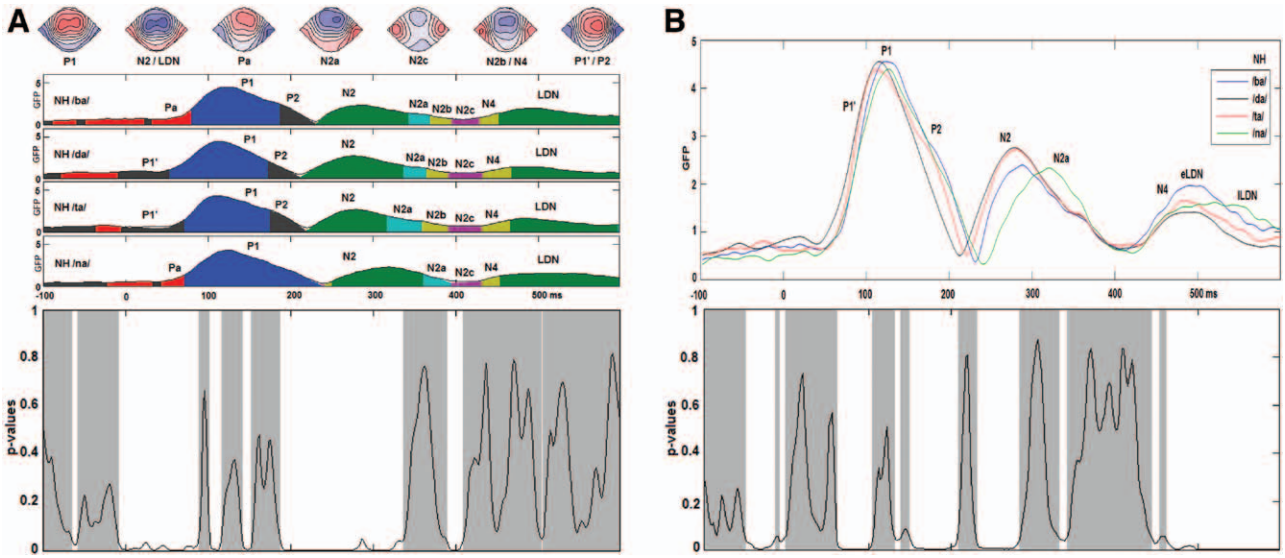

Fig. 1. Syllable effects in the normal-hearing (NH) group between −100 and 600 msec for the four different syllables (/ba/, /da/, /ta/, and /na/). A, In the first panel, all map topographies that we observed are labeled according to the relevant components. The second panel represents the four different syllable conditions in normal-hearing children indicating periods of stable topography (microstates) under the global field power (GFP) curves with a different color for each distinct component. All microstates are labeled with the corresponding components. The subsequent panel delineates *p* values with white columns for significant ($p < 0.05$) and gray columns for not significant time windows, as indicated by the topographical analysis of variance (TANOVA). B, In the first panel, GFP curves of normal-hearing children are presented with a different color for each syllable condition (blue /ba/, green /na/, black /da/, and red /ta/). Corresponding components are marked over the GFP amplitude curves, and significant syllable effects are indicated in the subsequent panel with white columns for significant ($p < 0.05$) and gray columns for nonsignificant time frames. LDN indicates late discriminative negativity.

In the prestimulus and early poststimulus intervals, there were a series of components with a frontal-midline positivity and negativities at temporal to occipital sites. Earlier studies have labeled these components as Pa and frontal and central subcomponents of P1 and P2 (Näätänen 1987; Ceponiene et al. 2008), which are typically evoked by a brief auditory stimulus. We found another early component similar to P1 that we have now labeled P1'. Later components were defined by positivities covering electrodes in the temporal lobe and parts of the occipital lobe and frontal negativities that have been labeled N2, N2a, N2b, N2c, N4, and the late discriminative negativity (LDN) according to the literature (Näätänen 1987; Sams & Näätänen 1991) (Fig. 1A). The N2a is generated by a difference between a sensory memory trace for a previously presented auditory stimulus and another incoming stimulus that violates the neural memory presentation of the recently heard stimulus. The N2b is elicited by a process representing an anticipated stimulus itself, whereas the N2c is evoked when various stimuli have to be divided up into different categories. The N2 and the LDN are closely linked to the evaluation of the incoming stimuli, and the N4 is further related to phonemic priming (Näätänen 1987).

Overall Statistical Analysis

A TANOVA of all four syllables showed significant effects across large time spans. To rule out the possibility of type 1 statistical error, we compared the amount of time points meeting the statistical criterion of $p < 0.05$ for the distribution of this count against the null-hypothesis, as determined through data randomization in a single overall test (Koenig & Melie-García 2010). We found the difference was highly significant ($p < 0.0002$), indicating that the overall null-hypothesis of no effects of syllables could be rejected and an in-depth analysis was thus conducted. The underlying differences between ERPs evoked by distinct syllables are discussed in the following sections as a function of the different aspects of speech sound discrimination. Based on the overall TANOVA and on the timing of the microstates, the following five time windows were defined and used for further comparisons: (1) a time window from −10 to 88 msec ($p = 0.001$), which was divided into two parts for further group comparisons (see below): an initial subinterval around stimulus presentation (−10 to 6 msec, $p = 0.004$) and a second shortly after stimulus presentation (6 to 88 msec, $p < 0.0001$); (2) time windows from 100 to 116 msec ($p = 0.007$) and (3) from 142 to 152 msec ($p = 0.004$); (4) a period from 188 to 336 msec ($p < 0.0001$) that was divided into subintervals from 188 to 214 msec ($p < 0.0001$) and from 216 to 336 msec ($p < 0.0001$); and (5) a time window from 390 to 408 msec ($p = 0.004$).

Furthermore, we tested for significant differences of the global amplitude that might be evoked by the four different syllables. As in the TANOVA, an overall test for the count of time points meeting the criterion of $p < 0.05$ was highly significant ($p < 0.0002$), and the null hypothesis of no difference between syllables was therefore rejected. We compared the effects of GFP (Fig. 1B) covering time spans (1) from −48 to −2 msec ($p = 0.006$), (2) from 62 to 104 msec ($p < 0.0001$) which was divided into two separate intervals (from 62 to 86 msec, $p = 0.001$; and from 88 to 104 msec, $p < 0.0001$); (3) a time period from 150 to 210 msec ($p < 0.0001$); (4) a time window from 232 to 282 msec ($p = 0.004$); and (5) a later window from 462 to 600 msec ($p < 0.0001$), which was separated into two periods (from 462 to 528 msec [$p = 0.001$] and the final window from 530 to 600 msec [$p < 0.0001$]).

Physiological Correlates of Spectral, Temporal, and Spectrotemporal Features Discrimination in NH Subjects • Spectral feature discrimination was defined by the difference between the syllables /da/ and /ba/ lying within the mid-frequency band. Based on the overall TANOVA result, during the P2 component from 188 to 214 msec, further comparison analysis showed a *t* map with a more left than right temporo-occipital positivity (t_{max} at P7 = 4.049) and a more left-weighted centroparietal negativity (t_{min} at C3 = −3.999) when comparing the syllable /da/ with /ba/ (Fig. 2A). A significant processing speed difference between microstates was detected for the P2 component offset (within 212 to 230 msec for all syllables but /na/, $p = 0.0224$), and no P2 component was present in the /na/ syllable condition. In addition, GFP analysis revealed that during the P2 component (150 to 210 msec, $p = 0.014$) and the LDN component (462 to 528 msec, $p = 0.001$), the /ba/ syllable had more field strength compared with all other syllables. The results indicate that the P2 component appears to be associated with spectral feature analysis, whereas early parts of the LDN component might be attributed to spectral control mechanisms.

Temporal feature discrimination was analyzed by the VOT distinction between the syllables /da/ and /ta/ comprising the high-frequency band. Based on the overall TANOVA, post hoc analysis of the P1' component for the time period from 6 to 88 msec revealed a *t* map (Fig. 2B) with more bilateral frontal positivity (t_{max} at F8 = 5.831) and more bilateral parietal negativity (t_{min} at P7 = −5.218) evoked by the syllables /da/ compared with /ta/. Furthermore, when comparing the syllable /da/ with /ta/ during the N2 component from 216 to 336 msec, the *t* map showed left parietal positivity (t_{max} at P3 = 3.800) and more right frontocentral negativity (t_{min} at Fz = −4.749) (Fig. 2C). Microstate latency effects occurred for the P1' offset ($p = 0.0497$) for /da/ and /ta/ at 54 and 72 msec and for the duration of P1' ($p = 0.0350$) with latencies for /da/ and /ta/ between 62 and 76 msec. Furthermore, a speed processing effect for the N2 onset ($p < 0.0001$) was found. The syllables /da/, /ta/, and /ba/ all had an onset within 214 to 232 msec, whereas /na/ began at 252 msec. This speed distinction between syllables was probably evoked by the different consonantal duration between short plosives and long-lasting nasals. GFP analysis of the first period of the P1' component (62 to 86 msec) ($p = 0.005$) indicated that the strongest GFP appeared for the syllable /da/ against all other syllables. During the subsequent interval (88 to 104 msec) ($p = 0.017$) and over the course of the N2 component (232 to 282 msec) ($p = 0.005$), the syllable /ta/ had a greater GFP compared with the other syllables. These results led us to the conclusions that the P1' component was involved in temporal feature analyses and the N2 component had functional importance for control mechanisms with respect to temporal feature discrimination.

Spectrotemporal feature discrimination was represented by the difference between the syllables /da/ and /na/, which covered the low-frequency and mid-frequency band. Post hoc analyses of the overall TANOVA in the P1 time window from 100 to 116 msec contrasting /da/ against /na/ showed a *t* map with a bilateral, but more right than left frontal positivity (t_{max} at Fpz = 4.128) and a bilateral centroparietal negativity (t_{min} at C3 = −2.735) (Fig. 2D). Comparisons between /ta/ and /ba/ during the central P1 from 142 to 152 msec generated a *t* map with bilateral frontal positivity (t_{max} at Fp2 = 3.359) and a centroparietal negativity (t_{min} at Pz = −3.149) (Fig. 2E). Significant temporal differences for the P1 offset ($p < 0.0001$) and duration

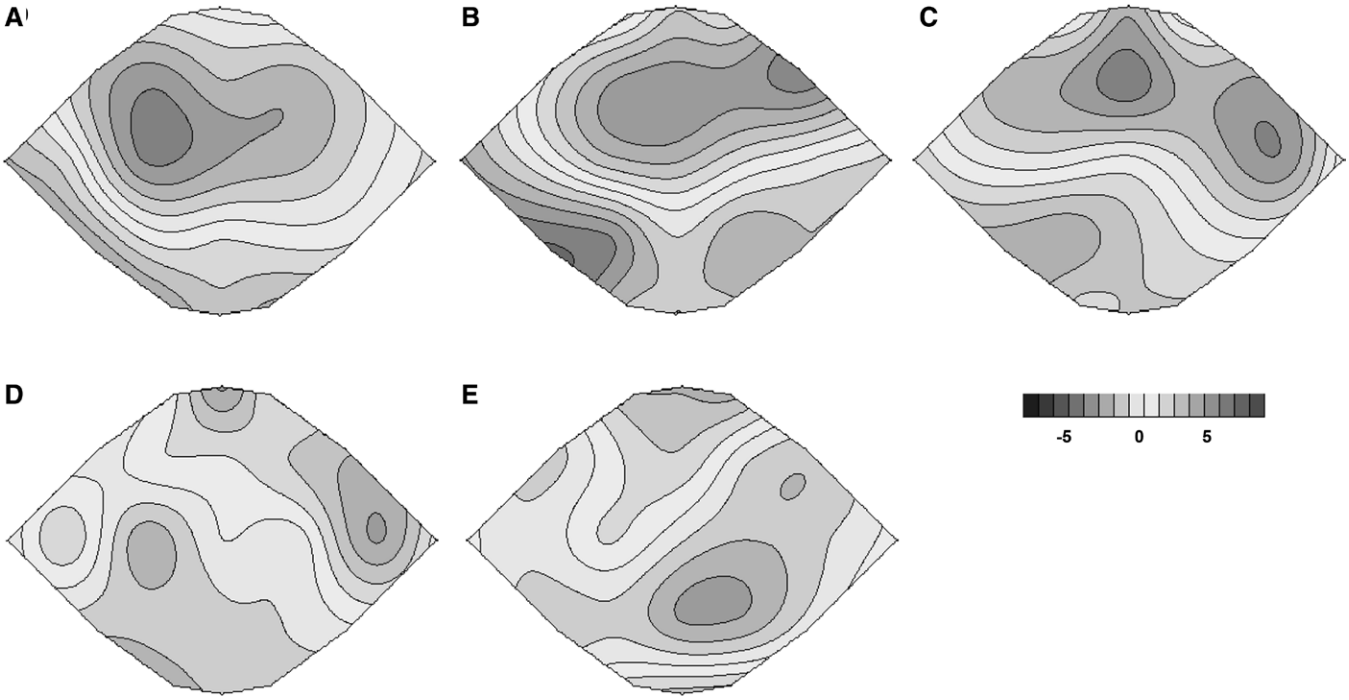

Fig. 2. A *t* map of syllable differences in the normal-hearing group. Differences between /da/ and /ba/ over the course of the P2 component (188 to 214 msec) (A), differences between /da/ and /ta/ over the course of the P1′ component (6 to 88 msec) (B), differences between /da/ and /ta/ over the course of the N2 component (216 to 336 msec) (C), differences between /da/ and /na/ over the course of the frontal P1 component (100 to 116 msec) (D), and differences between /ta/ and /ba/ over the course of the central P1 component (142 to 152 msec) (E). Contours are presented in steps of 1 *t*.

($p = 0.0138$) were observed. While the syllables /ba/, /da/, and /ta/ had their offset within 174 to 188 msec and their duration between 104 and 120 msec, /na/ ended after 238 msec with a duration of 166 msec. This difference was a manifestation of the distinct consonantal duration of the phoneme classes (plosiv against nasal). GFP analysis of the late LDN component (530 to 600 msec) ($p = 0.004$) revealed that the /na/ syllable exhibited more field strength compared with all other syllables. Taken together, the data indicate that the P1 and the late LDN component can be assigned to roles for spectrotemporal feature analysis and control mechanisms. Because the /da/ versus /na/ contrast covered spectrotemporal feature analyses from low-frequency to mid-frequency bands, the results further indicated that only the frontal P1 was involved in this process, whereas the central P1 was associated with the /ba/–/ta/ distinction, covering mid-frequency to high-frequency bands.

Group Comparisons Between NH and HI Subjects in Physiological Correlates of Spectral, Temporal, and Spectrotemporal Features Discrimination • The only group difference of the HI participants related to ERPs associated with spectral feature discrimination was found in a GFP syllable by group interaction at a late process stage from 504 to 522 msec ($p = 0.036$) corresponding to the time when the early LDN component appeared, when comparing /ba/ against the remaining syllables (Fig. 3A). This indicated that the HI group, when compared with the NH group, showed diminished control mechanisms with respect to spectral feature discrimination.

Group comparisons regarding temporal feature discrimination of /da/ versus /ta/ differences yielded no significant results in the TANOVA, but the GFP analysis revealed a main effect from 90 to 120 msec ($p = 0.020$) that overlapped with the P1' and the frontal P1 components, whereas the NH had a stronger global field strength (4.19 μV versus 3.22 μV). In addition, GFP results showed a syllables by group interaction from 228 to 242 msec ($p = 0.013$) during the presence of the N2 component. Visualization of the results (Fig. 3B) indicated that this was mainly due to a differential effect that appeared during the presentation of the syllable /ta/. Indeed, the NH group had the strongest GFP for /ta/ ($p = 0.020$), whereas no significant differences ($p = 0.541$) between the syllables were found in the HI group (Fig. 3B). These results led us to the conclusions that the HI group compared with the NH group exhibited less neuronal activity related to temporal feature analysis and control mechanisms of temporal feature discrimination.

No group effects could be found with respect to spectrotemporal feature discrimination, implying that NH and HI groups showed similar processing.

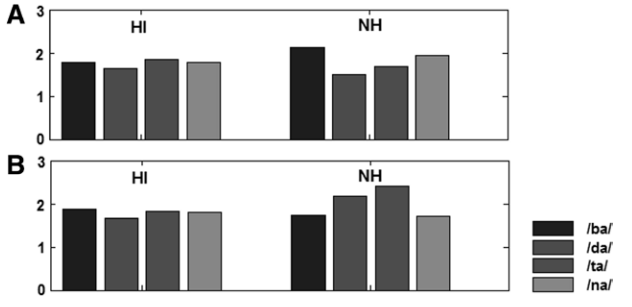

Fig. 3. Global field power interaction between the hearing-impaired (HI) and normal-hearing (NH) groups. Comparing /ba/ against the remaining syllables over the course of the late discriminative negativity component (504 to 522 msec) (A) and over the course of the N2 component (228 to 242 msec) (B).

Physiological Correlates of Distributional Statistical Learning in NH Subjects • In the behavioral discrimination task, the category with infrequent syllable pairs was taken as an indicator for distributional statistical learning. As for the ERPs, distributional statistical learning was quantified by the difference of the /da/-evoked ERPs compared with the remaining syllables, as indicated by the more frequent appearance of /da/ (50% versus 16.66%).

Post hoc analysis of the overall TANOVA demonstrated that for the time interval corresponding to −10 to 6 msec, subjects exhibited a frontocentral positivity (t_{max} at Cz = 2.723) and a left temporal negativity (t_{min} at T7 = −4.006) with the highest values between /da/ and /na/ (Fig. 4A). However, all other conditions possessed similar *t* maps with a right frontocentral positivity and a left temporo-occipital negativity, when compared with /da/ (Fig. 4B, C). In addition, children with NH abilities showed significantly lower GFP for /da/-evoked potentials in the prestimulus period from −48 to −2 msec ($p = 0.006$) (Fig. 4D). Further syllable comparisons produced a *t* map for the N2b component from 390 to 408 msec with the highest values between /da/ and /na/ with a maximal left temporal positivity (t_{max} at T7 = 5.914) and a centroparietal negativity (t_{min} at Pz = −4.215) (Fig. 4E), but all other conditions compared with /na/ possessed similar *t* maps. In summary, these two time frames can be assigned to roles for statistical learning mechanisms. The first time window around stimulus onset appears to be sensitive to frequency of occurrence, whereas the second time range seems to be attuned to slight acoustic changes within the presented stimuli.

Group Comparisons Between NH and HI Subjects in Physiological Correlates of Distributional Statistical Learning • When the HI group was tested for physiological differences in indices of statistical learning in the time period where the NH group showed an effect (−10 to 6 msec), there was a significant group interaction in the TANOVA ($p = 0.013$). Further post hoc group comparisons of the difference of /da/ against the remaining syllables yielded a *t* map with a right central positivity (t_{max} at C4 = 2.763) and a left temporal negativity (t_{min} at T7 = −3.327) (Fig. 4F). This *t* map of the group difference closely resembled the effect of distributional statistical learning itself in the NH group, thus indicating that in the early stage the HI group recruited less neural activity than was found to be associated with distributional statistical learning in the NH group. GFP analysis revealed no group effect ($p = 0.398$), that is, the HI group showed a similar reduction in GFP for the frequent syllable as compared with the NH group. In the second time frame where the NH group showed an effect (390 to 408 msec), there was a nonsignificant trend in the HI group (TANOVA; $p = 0.068$), implying that if there was a higher sample size a significant effect might be observed in this time window.

Physiological Results With Age Taken as a Covariate • A TANCOVA analysis implemented separately with the NH and the HI group revealed main effects covering time spans from 70 to 82 msec for the NH group ($p = 0.030$, Fig. 5A) and from 38 to 104 msec for the HI group ($p = 0.004$, Fig. 5B). The map topography of the age effects closely resembled the topography of the P1/P1' complex, indicating that the P1/P1' complex becomes stronger with age. When considered in combination with the observation that children with HL exhibited less GFP in the same time period, this suggests a maturational lag in children with HI.

Speech Production Effects

Behavioral Results of Consonantal Duration and Intensity • The consonantal duration of the speech production

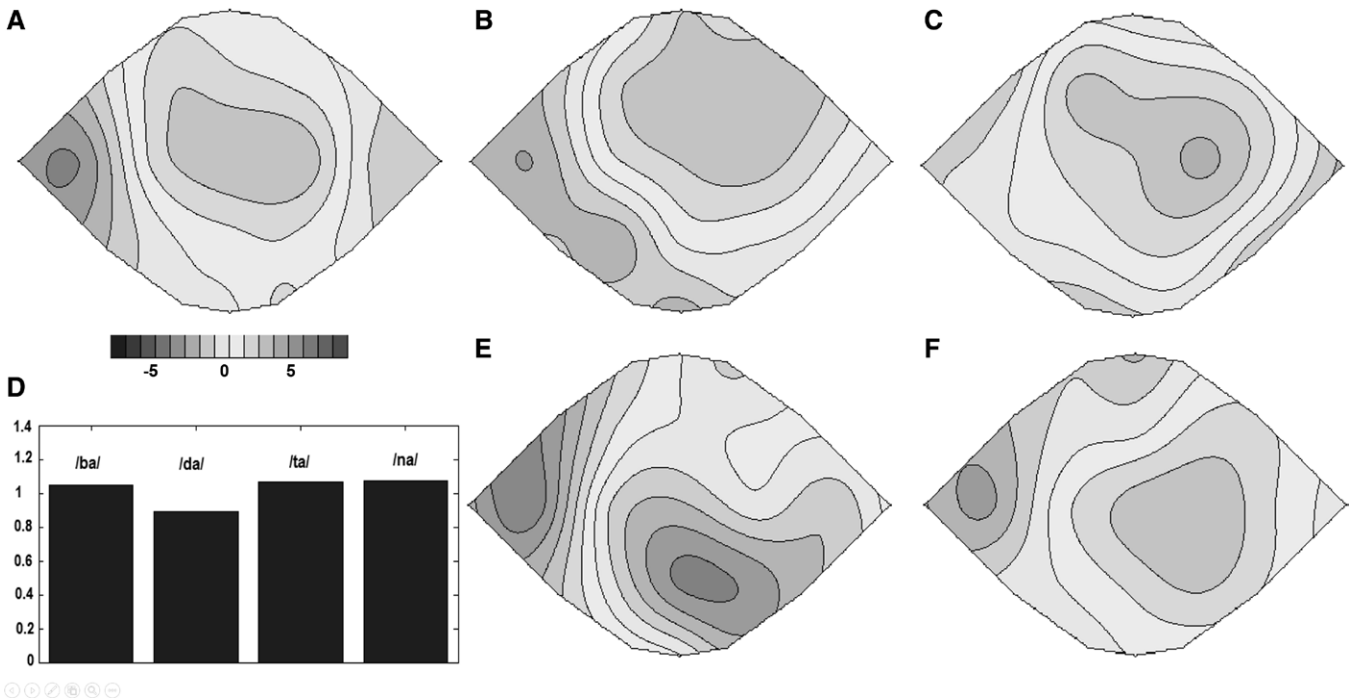

Fig. 4. Analysis of statistical learning in the normal-hearing (NH) group and between the hearing-impaired and NH groups. A *t* map of syllable differences between the standard stimulus /da/ and (A) the deviant /na/, (B) the deviant /ta/, (C) the deviant /ba/ in the NH group near stimulus onset (−10 to 6 msec), (D) global field power amplitude during the time window from −48 to −2 msec in the NH group, (E) a *t* map of syllable differences between /da/ and /na/ over the course of the N2b component (390 to 408 msec) in the NH group, and (F) NH and hearing-impaired groups difference (−10 to 6 msec) between /da/ and all remaining syllables. Contours are presented in steps of 1 *t*.

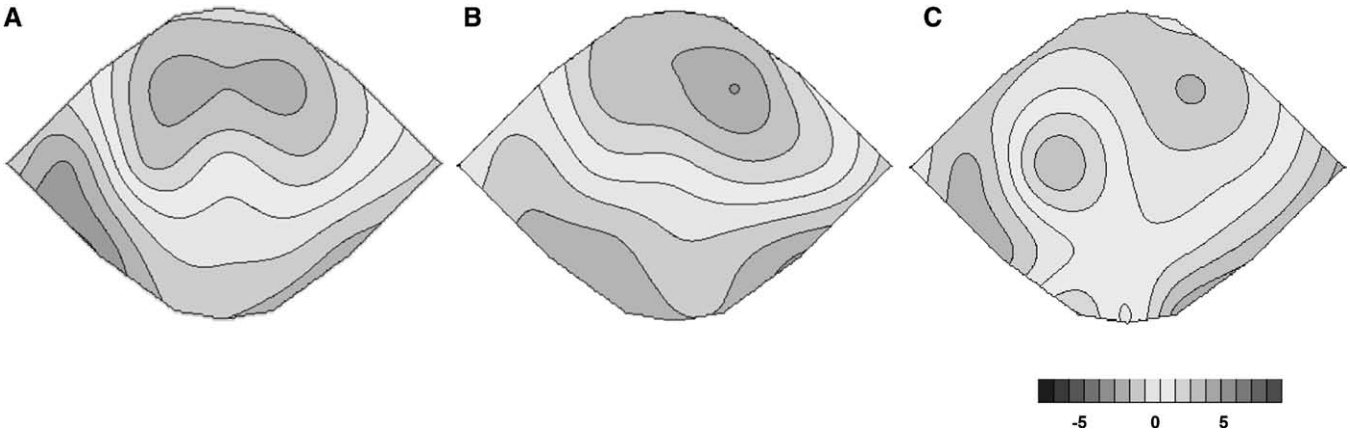

Fig. 5. Analysis of age and speech performance. The topographical analysis of covariance map of age during the P1′ component (70 to 82 msec) in the normal-hearing group (A), the topographical analysis of covariance map of age during the P1′ component (38 to 104 msec) in the hearing-impaired group (B), and *t* map of speech rate differences between the normal-hearing and hearing-impaired group during the time window from 150 to 168 msec (C). Contours are presented in steps of 1 *t*.

behavioral analysis revealed significant differences between hearing functions [$F(1, 920) = 24.26$, $p = 0.0005$ with monaural subjects producing a longer duration] and consonants [$F(5, 920) = 272.11$, $p = 0.0005$] with nasal producing a longer duration than plosives, and within the plosive phoneme class, the voiced was longer than the voiceless. On the other hand, no NH and HI groups effect [$F(1, 920) = 2.82$, $p = 0.1868$], no sex effect [$F(1, 920) = 4.44$, $p = 0.1059$], and no age effect [$F(1, 920) = 0.02$, $p = 0.8877$] could be found. In summary, consonantal speech duration played a key role in differentiating coordinative speech patterns; longer speech products correlated with an improved ability to produce distinctive duration patterns. Furthermore, no group effect with respect to the perseveratory precision of consonantal duration was observed, suggesting that differences in auditory perception do not affect durational speech production precision.

The intensity of the speech products behavioral analysis revealed significant differences between the NH and HI groups [$F(1, 920) = 7.26$, $p = 0.0171$ with HI producing louder speech], sex [$F(1, 920) = 7.68$, $p = 0.0171$, boys producing louder speech], age [$F(1, 920) = 121.18$, $p = 0.0005$, younger producing louder speech], and consonants [$F(5, 920) = 34.96$, $p = 0.0005$, short consonantal duration was louder]. However, no differences were observed between hearing functions [$F(1,920) = 1.72$, $p = 0.1896$]. Concerning the consonantal intensity, there appears to be a close relationship to speech duration. The shorter the speech products were the louder the speech intensities produced by the subjects. In contrast, children with HI were less capable in regulating speech intensity.

TABLE 5. Behavioral results of consonantal duration and intensity comparing group, sex, hearing functions, age, different consonants, and group effects in perseveratory precision

	Duration (*p*)	Intensity (*p*)
Groups	0.1868	0.0171
Sex	0.1059	0.0171
Hearing functions	0.0005	0.1896
Age	0.8877	0.0005
Consonants	0.0005	0.0005
Group effects in perseveratory precision	Not significant	0.0482 (syllable /ba/)

A significant NH and HI groups effect in perseveratory precision of consonantal intensity emerged only for the syllable /ba/ [$F(1, 24) = 4.34$, $p = 0.0482$]. This suggests that the attunement to intensity precision in short speech products is affected by auditory perception. The *p* values for multitesting were Bonferroni–Holm corrected. An overview of the behavioral speech production results is provided in Table 5.

Physiological Results With Maximal Speech Rate Taken as a Covariate • Since a *t* test evaluating maximal speech rate demonstrated no NH and HI groups effect ($t = 1.3341$, df = 24, $p = 0.1948$), the median was calculated to divide the expressive data into a group with higher and lower speech rates. Then ERP results revealed a TANCOVA main effect from 150 to 168 msec ($p = 0.026$). When comparing children with high versus low speech rates, the corresponding *t* map possessed a left centroparietal positivity (t_{max} at A2 = 3.523) and a left frontotemporal negativity (t_{min} at P7 = −2.343) (Fig. 5C). This result led us to conclude that children from the NH and HI groups with a fast speech rate showed more left-sided activity in sensorimotor cortices than children with a slow speech rate. However, the correlation between the maximal speech rate and the degree of HL in the HI group was not significant ($r = -0.06013928$, $p = 0.8453$), implying that velocity of speech production was not affected by the severity of HL.

DISCUSSION

In this study, we compared speech sound processing in NH preschool children to that in preschool children with peripheral HL. We focused on deciphering the spectral, temporal, and spectrotemporal features of speech sound discrimination. In addition, we elucidated features of distributional statistical learning and indications of a close connection between speech perception and production mechanisms. To clarify whether children with peripheral HL showed significant differences in their speech sound processing mechanisms, we carried out a comprehensive study that integrated a multifeature-stimulation paradigm, multichannel ERP analysis, and receptive and expressive behavioral tasks. Our approach allowed us to discriminate between the spectral, temporal, and spectrotemporal features of speech processing, thus allowing us to associate neural activity evoked by distinct speech components with their functional roles, and we further found evidence for distributional statistical learning in children with NH.

Integration With Previous Studies of Speech Perception

Our findings are in line with previous studies describing predictable Pa, P1, P2, N2, N2a, N2b, N2c, N4, and LDN components (Näätänen 1987; Maurer et al. 2003; Ceponiene et al. 2005; Ceponiene et al. 2008; Sussman et al. 2008). In addition, we found a component similar to P1, but in the preceding interval (6 to 88 msec) with bilateral frontal positivity and bilateral parietal negativity that we labeled P1'.

Based on association of ERP components with variations of specific features of the stimuli, our results indicated that the component after P1' was involved in temporal feature analyses. Building on the work by Ceponiene et al. (2008), who investigated only spectral feature differences, we found evidence not only for spectral feature processing during the two P1 subcomponents but also for spectrotemporal feature detection from low- to middle-frequency ranges in the first subcomponent and from middle- to high-frequency bands in the subsequent subcomponent. Furthermore, our results revealed that the P2 component has functional significance in spectral feature analyses, whereas late components such as the N2 (temporal) and the LDN (early: spectral; late: spectrotemporal) appeared to be involved in control mechanisms, as reported in the literature (Ceponiene et al. 2005; Ceponiene et al. 2008).

When comparing ERPs of NH and preschool children with HI, our findings showed significant differences in two relevant domains. First, children with peripheral HL exhibited less neural activity at disposition with respect to temporal feature analysis, as detected by a lower GFP during the presence of the P1' component. Based on the analysis of age effects, this difference may be attributed to a maturational lag between children with NH and children with HI. Second, children with HI demonstrated reduced GFP in the N2 component and the early LDN component. An abnormal N2 amplitude was previously reported by Koravand et al. (2013). These neural activity alternations lead to decreased control mechanisms over temporal and spectral features, which may provide an explanation for abnormal speech sound perception and control processes (Grant et al. 2007; Jerger 2007) and for disabilities of auditory temporal ordering of language-specific stimuli in these patients (Koravand & Jutras 2013).

Integration With Previous Studies of Speech Production

One of our hypotheses was that there is a close connection between neurophysiological speech sound perception and speech production processes. With respect to the aforementioned group differences in speech sound perception processing, this would indicate that speech sound production should also reveal group effects. However, our data did not support this possibility, as we found few group differences. On the other hand, previously reported effects of syllable features in normally hearing children could be well replicated. Our data revealed that children with NH abilities produced distinctive duration patterns between and within different phoneme classes, which has been previously reported in the literature (Goffman & Smith 1999; Smith 2006). With respect to consonantal duration, no differences between age groups were observed. In addition, our findings indicated a close relation between consonantal duration and the ability to regulate the intensity of speech products; shorter speech sound durations correlated with the louder implementation of consonantal realization in children with NH. In contrast to the duration of speech sounds, a significant age-related difference could be found in consonantal intensity. These results led us to the conclusions that the ability to simultaneously coordinate the duration of articulatory movements with the fine tuning of speech intensity varies with age.

Only a subgroup of children with peripheral HL, namely those with unilateral HL, showed a significant difference in consonant duration; we found that they produce consonants with longer duration values than do children with balanced binaural abilities, similar to the findings by Higgins et al. (2005). We found no group difference of perseveratory movements of consonant duration, in agreement with Baum and Waldstein (1991).

Group effects were observed for tests of the regulation of consonant intensity and perseveratory articulation movements of the voiced bilabial plosive. Children with HI differed only in a consonant with short speech sound duration. Koopmans-Van Beinum et al. (2001) reported a more generalized lack of coordination with respect to speech sound products in infants with HI. Based on these observations, we postulate that speech motor coordination undergoes maturational processes in infants and toddlers, such that at the age range from 4 to 7 years only subtle distinctions can be observed. Based on these behavioral findings, we assume that a peripheral impairment of auditory capacities has a decreasing impact on coordinating speech production skills with increasing age.

An additional novelty of this study was the use of behavioral results such as the maximal speech rate as a covariate in TANCOVAs. Interestingly, higher speech rates were associated with a more left-sided activity in sensorimotor cortices around 150 msec, a time span where spectral feature analysis was implemented in speech sound perception processing. In light of this result, we suppose that spectral feature analysis has an impact on the rate of syllable repetition in preschool children.

Integration With Previous Studies of Statistical Learning Mechanisms

According to the literature (Daltrozzo & Conway 2014), statistical learning consists of at least two different mechanisms containing a bottom-up implicit mechanism and a more attention-dependent top-down explicit mechanism. The first mechanism develops early in life and is sufficient for rudimentary language-related abilities, whereas the latter requires explicit knowledge of abstract rules necessary to improve language abilities. Regarding indications of distributional statistical learning mechanisms in the present study, ERP responses of NH children yielded two variant components. The first component arose around stimulus onset with a right frontocentral positivity and a left temporo-occipital negativity, while the other appeared around 400 msec in the N2b component. The first component was reflective of a preparatory phase that was sensitive to distributional input associated with implicit learning, whereas the second more anticipatory component was responsive to feature-specific distinctions between the presented stimuli and linked to explicit learning. These results are in agreement with previous studies describing preparatory and more attention-dependent components with respect to statistical learning (Stadler et al. 2006; Schröger et al. 2007).

When comparing ERPs between the NH and HI groups, children with HI showed diminished implicit statistical learning mechanisms, as indicated by the ERP-evoked difference

around stimulus onset, and a nonsignificant trend in the N2b component, which reflects explicit learning. These findings of diminished statistical learning mechanisms may explain why HI children show impairments in phonological short-term memory; phonological discrimination and awareness (Briscoe et al. 2001); articulation rate, which has been linked to verbal short-term memory processes (Stiles et al. 2012); and cross-modal learning of sequences (Conway et al. 2009). These verbal and nonverbal functions require the ability to create categories, which in turn depends on the frequency of occurrence of a given input.

Clinical Implications and Future Directions

An unexpected finding for both groups and age ranges was that all subjects discriminated syllables in the behavioral task at chance level. Furthermore, discrimination tasks revealed no group differences, while ERP findings did. These findings raise the question of whether alternative, more sensitive behavioral tasks than we used in this study would find effects for these phenomena.

Our work may have important clinical implications, as valid knowledge of speech sound discrimination abilities based on ERP in preschool children with HI can be applied in devising more optimal therapeutic strategies. Our findings emphasize the clinical importance of exerting language-specific auditory discrimination training that is focused on spectral and temporal distinctions. Effort should be made to monitor the appropriate frequency of occurrence of auditory input, characterize regularities of phonological processing to create phoneme classes, and increase awareness of regularities of not only auditory input but also incoming cross-modal information.

A limitation of our study is that our ERP findings allow conclusions to be made only for preschool children from ages 4 to 7 years. We observed a nonsignificant trend for the anticipatory N2b component, which may be due to the small sample size. To clarify this, our study needs to be replicated with more participants. Further studies are needed to clarify the divergence between ERP and behavioral findings. Specifically, the behavioral discrimination tasks should be implemented with language-specific stimuli, which are sensitive to slighter acoustic changes. Furthermore, studies focusing on children with unilateral HL, children using cochlear implant, or toddlers with HIs are important avenues for further research. Such studies will provide better methods for speech therapy before preschool age.

CONCLUSIONS

We found that speech sound perception is based—apart from spectral, temporal, and spectrotemporal feature analysis, and control mechanisms—on implicit and explicit statistical learning mechanisms. Children with peripheral HL showed delayed temporal feature analysis, reduced spectral and temporal control mechanisms, and diminished statistical learning mechanisms. Receptive deficits in children with HI resulted only in subtle abnormalities in speech production, reflected in a delay in acquiring the ability to simultaneously coordinate the fine tuning of the duration and intensity of articulatory movements. These findings indicate that the integrated perceiving-acting system is less affected in 4- to 7-year-old children with HL compared with statistical learning mechanisms.

ACKNOWLEDGMENTS

The authors thank all children and their parents for participating in this study, and the Department of Psychiatric Neurophysiology of the University Hospital of Clinical Psychiatry in Bern for use of their facilities and for their generous support.

This work was supported by the Department of General Linguistics of the University of Bern (financial support for the local Ethics Committee of Bern). T.K. is currently receiving a grant (no. 136249) from the Swiss National Science Foundation.

E.S.-E. recruited all subjects; designed and performed measurements; analyzed and interpreted data; and wrote the article. F.S.-E. analyzed data and provided critical revision. T.K. supervised experimental setup, analyzed data, and wrote the article.

The authors declare no other conflict of interest.

Address for correspondence: Esther Studer-Eichenberger, Lerchenweg 2, CH-3012 Bern, Switzerland. E-mail: studer_esther@bluewin.ch

Received September 12, 2014; accepted June 5, 2015.

REFERENCES

Aichert, I., Marquardt, C., Ziegler, W. (2005). Frequenzen sublexikalischer Einheiten des Deutschen: CELEX-basierte Datenbanken. *Neurolinguistik*, *19*, 55–81.

Aslin, R. N., Saffran, J. R., Newport, E. L. (1999). Statistical learning in linguistic and nonlinguistic domains. In B. MacWhinney (Ed.), *The Emergence of Language* (pp. 359–380). Mahwah, NJ; London, UK: Lawrence Erlbaum Associates, Publishers.

Bamford, J., & Saunders, E. (1985). *Hearing Impairment, Auditory Perception and Language Disability*. London, UK: Edward Arnold Ltd.

Baum, S. R., & Waldstein, R. S. (1991). Perseveratory coarticulation in the speech of profoundly hearing-impaired and normally hearing children. *J Speech Hear Res*, *34*, 1286–1292.

Bertoncini, J. (1993). Infants' perception of speech units: Primary representation capacities. In B. de Boysson-Bardies, S. de Schonen, P. W. Jusczyk, *et al*. (Eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life* (Vol. 69, pp. 249–257). Dordrecht, The Netherlands; Boston, MA; London, UK: Kluwer Academic Publishers.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words* (pp. 167–224). Cambridge, MA; London, UK: A Bradford Book The MIT Press.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience. Issues in Cross-Language Research* (pp. 171–204). Baltimore, MD: York Press.

Best, K.-H. (2004/2005). Laut- und Phonemhäufigkeiten im Deutschen. In *Göttinger Beiträge zur Sprachwissenschaft* (Vol. 10/11, pp. 21–31). Göttingen: Peust & Gutschmidt Verlag.

Boersma, P., & Weenik, D. (2009). *Praat (Version 5.1)*. Amsterdam: PRAAT. Phonetic Sciences, University of Amsterdam.

Briscoe, J., Bishop, D. V., Norbury, C. F. (2001). Phonological processing, language, and literacy: A comparison of children with mild-to-moderate-sensorineural hearing loss and those with specific language impairment. *J Child Psychol Psychiatry, 42*, 329–340.

Ceponiene, R., Alku, P., Westerfield, M., et al. (2005). ERPs differentiate syllable and nonphonetic sound processing in children and adults. *Psychophysiology*, *42*, 391–406.

Ceponiene, R., Torki, M., Alku, P., et al. (2008). Event-related potentials reflect spectral differences in speech and non-speech stimuli in children and adults. *Clin Neurophysiol*, *119*, 1560–1577.

Conway, C. M., Pisoni, D. B., Kronenberger, W. G. (2009). The importance of sound for cognitive sequencing abilities: The auditory scaffolding hypothesis. *Curr Dir Psychol Sci*, *18*, 275–279.

Council on Physical Therapy, American Medical Association (CPT-AMA). (1942). Tentative standard procedure for evaluation of the percentage of useful hearing loss in medico-legal cases. *JAMA*, *119*, 1108–1109.

Daltrozzo, J., & Conway, C. M. (2014). Neurocognitive mechanisms of statistical-sequential learning: What do event-related potentials tell us? *Front Hum Neurosci*, *8*, 437.

Gilley, P. M., Sharma, A., Dorman, M., et al. (2005). Developmental changes in refractoriness of the cortical auditory evoked potential. *Clin Neurophysiol*, *116*, 648–657.

Goffman, L., & Smith, A. (1999). Development and phonetic differentiation of speech movement patterns. *J Exp Psychol Hum Percept Perform*, *25*, 649–660.

Grant, K. W., Tufts, J. B., Greenberg, S. (2007). Integration efficiency for speech perception within and across sensory modalities by normal-hearing and hearing-impaired individuals. *J Acoust Soc Am*, *121*, 1164–1176.

Hall, T. A. (2007). Segmental features. In P. de Lacy (Ed.), *The Cambridge Handbook of Phonology* (pp. 311–334). Cambridge, UK: Cambridge University Press.

Higgins, M. B., McCleary, E. A., Ide-Helvie, D. L., et al. (2005). Speech and voice physiology of children who are hard of hearing. *Ear Hear*, *26*, 546–558.

Jasper, H. H. (1958). Report of the committee on methods of clinical examination in electroencephalography. *Electroencephalogr Clin Neurophysiol*, *35*, 370–375.

Jerger, S. (2007). Current state of knowledge: Perceptual processing by children with hearing impairment. *Ear Hear*, *28*, 754–765.

Johnson, K. (2012). *Acoustic and Auditory Phonetics* (3rd ed.). Malden, MA; Oxford, UK: Wiley-Blackwell.

Jung, T. P., Makeig, S., Westerfield, M., et al. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin Neurophysiol*, *111*, 1745–1758.

King, K. A., Campbell, J., Sharma, A., et al. (2008). The representation of voice onset time in the cortical auditory evoked potentials of young children. *Clin Neurophysiol*, *119*, 2855–2861.

Koenig, T., Kottlow, M., Stein, M., et al. (2011). RAGU: A free tool for the analysis of EEG and MEG event-related scalp field data using global randomization statistics. *Comput Intell Neurosci*, *2011*, 938925.

Koenig, T., & Melie-García, L. (2010). A method to determine the presence of averaged event-related fields using randomization tests. *Brain Topogr*, *23*, 233–242.

Koenig, T., Melie-García, L., Stein, M., et al. (2008). Establishing correlations of scalp field maps with other experimental variables using covariance analysis and resampling methods. *Clin Neurophysiol*, *119*, 1262–1270.

Koopmans-Van Beinum, F. J., Van den Dikkenberg-Pot, C. J. C., Van den Dikkenberg-Pot, I. (2001). Babbling and the lack of auditory speech perception: A matter of coordination? *Dev Sci*, *44*, 61–70.

Koravand, A., & Jutras, B. (2013). Auditory temporal-organization abilities in school-age children with peripheral hearing loss. *J Speech Lang Hear Res*, *56*, 1065–1074.

Koravand, A., Jutras, B., Lassonde, M. (2013). Auditory event related potentials in children with peripheral hearing loss. *Clin Neurophysiol*, *124*, 1439–1447.

Korczak, P. A., Kurtzberg, D., Stapells, D. R. (2005). Effects of sensorineural hearing loss and personal hearing AIDS on cortical event-related potential and behavioral measures of speech-sound processing. *Ear Hear*, *26*, 165–185.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., et al. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philos Trans R Soc Lond B Biol Sci*, *363*, 979–1000.

Kujala, T., Kallio, J., Tervaniemi, M., et al. (2001). The mismatch negativity as an index of temporal processing in audition. *Clin Neurophysiol*, *112*, 1712–1719.

Ladefoged, P. (2003). *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques*. Malden, MA; Oxford, UK; Carlton: Blackwell Publishing Ltd.

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalogr Clin Neurophysiol*, *48*, 609–621.

Mahajan, Y., & McArthur, G. (2012). Maturation of auditory event-related potentials across adolescence. *Hear Res*, *294*, 82–94.

Maurer, U., Bucher, K., Brem, S., et al. (2003). Development of the automatic mismatch response: From frontal positivity in kindergarten children to the mismatch negativity. *Clin Neurophysiol*, *114*, 808–817.

Maye, J., Weiss, D. J., Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Dev Sci*, *11*, 122–134.

Maye, J., Werker, J. F., Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101–B111.

Michel, C. M., Koenig, T., Brandeis, D. (2009). Electrical neuroimaging in the time domain. In C. M. Michel, T. Koenig, D. Brandeis, et al. (Eds.), *Electrical Neuroimaging* (pp. 111–143). Cambridge, UK: Cambridge University Press.

Murray, M. M., Brunet, D., Michel, C. M. (2008). Topographic ERP analyses: A step-by-step tutorial review. *Brain Topogr*, *20*, 249–264.

Näätänen, R. (1987). Event-related brain potentials in research of cognitive processes—A classification of components. In E. Van der Meer & J. Hoffmann (Eds.), *Knowledge Aided Information Processing* (pp. 241–273). Amsterdam, The Netherlands: Elsevier Science Publishers B. V.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, *38*, 1–21.

Näätänen, R., Kujala, T., Escera, C., et al. (2012). The mismatch negativity (MMN)—A unique window to disturbed central auditory processing in ageing and different clinical conditions. *Clin Neurophysiol*, *123*, 424–458.

Näätänen, R., Pakarinen, S., Rinne, T., et al. (2004). The mismatch negativity (MMN): Towards the optimal paradigm. *Clin Neurophysiol*, *115*, 140–144.

Nittrouer, S. (1996). Discriminability and perceptual weighting of some acoustic cues to speech perception by 3-year-olds. *J Speech Hear Res*, *39*, 278–297.

Oates, P. A., Kurtzberg, D., Stapells, D. R. (2002). Effects of sensorineural hearing loss on cortical event-related potential and behavioral measures of speech-sound processing. *Ear Hear*, *23*, 399–415.

Paterson, M. M. (1994). Articulation and phonological disorders in hearing-impaired school-aged children with severe and profound sensorineural losses. In J. E. Bernthal & N. W. Bankson (Eds.), *Child Phonology: Characteristics, Assessment and Intervention with Special Populations* (pp. 199–224). New York, NY: Georg Thieme Verlag.

Pelucchi, B., Hay, J. F., Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Dev*, *80*, 674–685.

Pompino-Marschall, B. (2009). *Einführung in die Phonetik* (3rd ed.). Berlin, Germany: Walter de Gruyter.

R-Development-Core-Team. (2010). *R: A Language and Environment for Statistical Computing (Version 2.11.1)*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved May 31, 2010, from http://www.R-project.org.

Reetz, H., & Jongman, A. (2009). *Phonetics, Transcription, Production, Acoustics, and Perception* (*Vol. 22*). Malden, MA; Oxford, UK; West Sussex, UK: Wiley-Blackwell.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Curr Dir Psychol Sci*, *12*, 110–114.

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Dev Psychol*, *39*, 484–494.

Sams, M., & Näätänen, R. (1991). Neuromagnetic responses of the human auditory cortex to short frequency glides. *Neurosci Lett*, *121*, 43–46.

Schneider, W., Eschman, A., Zuccolotto, A. (2001). *E-Prime user's guide.* Pittsburgh, PA: Psychology Software Tools, Inc.

Schröger, E., Bendixen, A., Trujillo-Barreto, N. J., et al. (2007). Processing of abstract rule violations in audition. *PLoS One*, *2*, e1131.

Singh, S., Liasis, A., Rajput, K., et al. (2004). Event-related potentials in pediatric cochlear implant patients. *Ear Hear*, *25*, 598–610.

Smith, A. (2006). Speech motor development: Integrating muscles, movements, and linguistic units. *J Commun Disord*, *39*, 331–349.

Stadler, W., Klimesch, W., Pouthas, V., et al. (2006). Differential effects of the stimulus sequence on CNV and P300. *Brain Res*, *1123*, 157–167.

Stevens, K. N. (1998). *Acoustic Phonetics*. Cambridge, UK: Massachusetts Institute of Technology.

Stiles, D. J., McGregor, K. K., Bentler, R. A. (2012). Vocabulary and working memory in children fit with hearing aids. *J Speech Lang Hear Res*, *55*, 154–167.

Strik, W. K., Fallgatter, A. J., Brandeis, D., et al. (1998). Three-dimensional tomography of event-related potentials during response inhibition: Evidence for phasic frontal lobe activation. *Electroencephalogr Clin Neurophysiol*, *108*, 406–413.

Sussman, E., Steinschneider, M., Gumenyuk, V., et al. (2008). The maturation of human evoked brain potentials to sounds presented at different stimulus rates. *Hear Res*, *236*, 61–79.

Thiessen, E. D., Kronstein, A. T., Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychol Bull*, *139*, 792–814.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Dev Psychol*, *39*, 706–716.

Thoonen, G., Maassen, B., Gabreëls, F., et al. (1999). Validity of maximum performance tasks to diagnose motor speech disorders in children. *Clin Linguist Phon*, *13*, 1–23.

Thoonen, G., Maassen, B., Wit, J., et al. (1996). The integrated use of maximum performance tasks in differential diagnostic evaluations among children with motor speech disorders. *Clin Linguist Phon*, *10*, 311–336.

Tyler, M. D., Best, C. T., Goldstein, L. M., et al. (2014). Investigating the role of articulatory organs and perceptual assimilation of native and non-native fricative place contrasts. *Dev Psychobiol*, *56*, 210–227.

Werker, J. F., Pons, F., Dietrich, C., et al. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*, *103*, 147–162.

Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Dev Psychobiol*, *46*, 233–251.

Wunderlich, J. L., Cone-Wesson, B. K., Shepherd, R. (2006). Maturation of the cortical auditory evoked potential in infants and young children. *Hear Res*, *212*, 185–202.

REFERENCE NOTE

1. According to analyzed data of the SNF-project 100011–116271/1: "Quantitative Ansätze zu einer Sprachgeographie der schweizerdeutschen Prosodie." Institute of Linguistics, University of Bern, Switzerland, 2005–2008.